



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C12N 15/12, C07K 15/06 C12Q 1/68	A2	(11) International Publication Number: WO 94/00572 (43) International Publication Date: 6 January 1994 (06.01.94)
--	----	--

(21) International Application Number: PCT/US93/05794

(22) International Filing Date: 18 June 1993 (18.06.93)

(30) Priority data:
07/903,466 22 June 1992 (22.06.92) US

(71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 300 Lakeside Drive, 22nd Floor, Oakland, CA 94612-3550 (US).

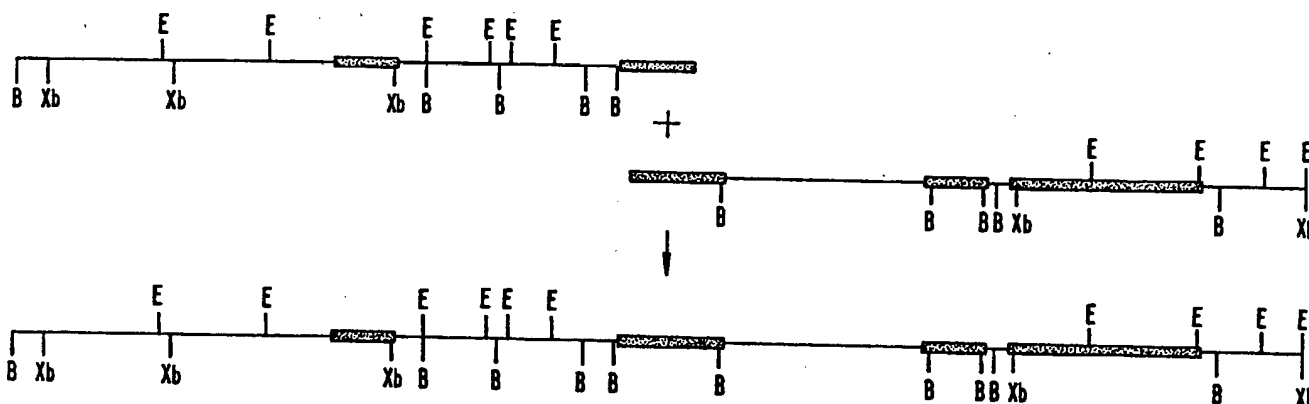
(72) Inventors: MURNANE, John, P. ; 1150 Stanyan Street, San Francisco, CA 94117 (US). PAINTER, Robert, B. ; 2933 Dolores Way, Burlingame, CA 94010 (US). KAPP, Leon, N. ; 49 Point San Pedro Road, San Rafael, CA 94901 (US). YU, Loh-Chung ; 1917 Alameda De las Pulgas, Redwood City, CA 94061 (US).

(74) Agent: LAUDER, Leona, L.; 177 Post Street, Suite 800, San Francisco, CA 94108-4731 (US).

(81) Designated States: AT, AU, BB, BG, BR, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published*Without international search report and to be republished upon receipt of that report.*

(54) Title: GENE FOR ATAXIA-TELANGIECTASIA COMPLEMENTATION GROUP D (ATDC)



(57) Abstract

Disclosed herein is an AT gene for complementation group D, the *ATDC* gene and fragments thereof. Nucleic acid probes for said gene are provided as well as proteins encoded by said gene, cDNA therefrom, preferably a 3 kilobase (kb) cDNA, and recombinant nucleic acid molecules for expression of said proteins. Further disclosed are methods to detect mutations in said gene, preferably methods employing the polymerase chain reaction (PCR).

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NE	Niger
BE	Belgium	GN	Guinea	NL	Netherlands
BF	Burkina Faso	GR	Greece	NO	Norway
BG	Bulgaria	HU	Hungary	NZ	New Zealand
BJ	Benin	IE	Ireland	PL	Poland
BR	Brazil	IT	Italy	PT	Portugal
BY	Belarus	JP	Japan	RO	Romania
CA	Canada	KP	Democratic People's Republic of Korea	RU	Russian Federation
CF	Central African Republic	KR	Republic of Korea	SD	Sudan
CG	Congo	KZ	Kazakhstan	SE	Sweden
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovak Republic
CM	Cameroon	LU	Luxembourg	SN	Senegal
CN	China	LV	Latvia	TD	Chad
CS	Czechoslovakia	MC	Monaco	TG	Togo
CZ	Czech Republic	MG	Madagascar	UA	Ukraine
DE	Germany	ML	Mali	US	United States of America
DK	Denmark	MN	Mongolia	UZ	Uzbekistan
ES	Spain			VN	Viet Nam
FI	Finland				

- 1 -

GENE FOR ATAXIA-TELANGIECTASIA
COMPLEMENTATION GROUP D (ATDC)

FIELD OF THE INVENTION

The present invention is in the general area of medical genetics. More specifically, it relates to the identification of a new gene--the ataxia-telangiectasia group D gene--the ATDC gene.

This invention was made with Government support under Contract Nos. DE-AC03-76-SF01012 and W-7405-ENG-48, awarded by the Department of Energy. The Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

Ataxia-telangiectasia (AT) is a human autosomal recessive disease that exhibits progressive neuromuscular problems, immunodeficiencies, a high incidence of lymphoreticular cancer, and sensitivity to ionizing radiation [Taylor, "Cytogenetics of ataxia telangiectasia," IN: Bridges and Harnden (eds.) Ataxia-telangiectasia--a cellular and molecular link between cancer, neuropathology and immune deficiency, pp. 53-82 (Wiley, Chichester 1988); Boder, "Ataxia telangiectasia an overview," IN: Gatti and Swift (eds.), Ataxia telangiectasia: genetics, neuropathology and immunology of a degenerative disease of childhood, pp. 1-63 (Alan R. Liss; New York, 1985); and Morrell et al., J. Natl. Cancer Inst., 77: 89-92 (1986)]. AT heterozygotes, which constitute as much as 3% of the human population, have been reported to have an increased risk of cancer after exposure to ionizing radiation.

Cells from patients with AT display two hallmark characteristics: hypersensitivity to the killing effects of ionizing radiation [Taylor et al., Nature, 258: 427-429 (1975)] and resistance to the inhibiting effects of ionizing radiation on the rate of DNA synthesis, that is,

- 2 -

radioresistant DNA synthesis [Young and Painter, Hum. Genet., 82: 113-117 (1989)]. Thus, the identification of genes that are responsible for those abnormalities would greatly further the understanding of human radiosensitivity and the regulation of DNA replication after radiation-induced DNA damage. The characteristic of AT cells to exhibit radioresistant DNA synthesis has been used to establish the presence of several complementation groups within this disease [Jaspers and Bootsma, PNAS (USA), 79: 2641-2644 (1982); Murnane and Painter, PNAS (USA), 79: 1960-1963 (1982); Jaspers et al., Cytogenet. Cell Genet., 49: 259-263 (1988)].

Despite extensive investigation, the underlying defects responsible for the pleiotropic abnormalities presented by AT remain unknown. Genetic linkage analysis [Gatti et al., Nature, 336: 577-580 (1988); McConville et al., Nucl. Acids Res., 18: 4335-4343 (1990a); McConville et al., Hum. Genet., 85: 215-220 (1990b); Sanal et al., Am. J. Hum. Genet., 47: 860-866 (1990); and Ziv et al., Genomics, 9: 373-375 (1991)] and chromosome transfer studies [Lambert et al., PNAS (USA), 88: 5907-5911 (1991)] have shown that the gene(s) associated with three complementation groups are all located at the chromosomal region 11q22-q23.

Complementation groups A (AT-A) and C (AT-C) have been mapped by genetic linkage analysis [Gatti et al. 1988; McConville et al. 1990a and 1990b; Sanal et al. 1990; Ziv et al. 1991; and Foroud et al., Am. J. Hum. Genet., 49: 1263-1279 (1991)]. Using families from mixed complementation groups, two groups of investigators independently reported linkage of the AT gene(s) to two separate regions, one of which is near THY1 [McConville et al. 1990a and 1990b; Sanal et al. 1990; Foroud et al. 1991]. It was subsequently determined that the genes for AT-A and AT-C are located within the more centromeric of these two regions [McConville et al. 1990a; Sanal et al. 1990; Foroud et al. 1991], although it was concluded that the AT gene in a small subset of families could map to the second locus near THY1 [Sanal et al. 1990]. A study by Gatti et al. [in paper entitled "Ataxia-telangiectasia: linking evidence for genetic heterogeneity,"

- 3 -

presented at AACR (American Association for Cancer Research) Special Conference "Cellular Responses to Environmental DNA Damage" in Banff, Alberta (Canada) December 1-6, 1991] which excluded families in complementation groups A and C, concluded
5 that a gene for an additional complementation group does show linkage to the region near THY1. The gene for complementation group D (AT-D) was a likely candidate for linkage near THY1 as complementation group D is the next most common
10 complementation group for AT after A and C [Jaspers et al. 1988].

Functional complementation has been used to prove the identity of several genes that provide resistance to various DNA-damaging agents [van Duin et al., Cell, 44: 913-923 (1986); Thompson et al., Mol. Cell Biol., 10: 6160-6171
15 (1990); and Weeda et al., Mol. Cell Biol., 10: 2570-2581 (1990)].

Using the sensitivity of AT cells to ionizing radiation, Kapp and Painter [Int. J. Radiat. Biol., 56: 667-675 (1989)] attempted to complement the defect in an AT cell
20 line (AT5BIVA) from complementation group D (AT-D) by transfection with a human cosmid library containing a selectable neo gene. The combined selection by ionizing radiation and G418 resulted in the isolation of an AT cell line (1B3) that is partially resistant (approximately 50% of
25 normal) to ionizing radiation and produces fewer radiation-induced chromosome aberrations, but retains the AT characteristic of radioresistant DNA synthesis. Southern blot analysis demonstrated that the 1B3 cell line contains at least three cosmids that appear to be integrated in tandem and
30 coamplified [Kapp and Painter 1989, id.]. Transfer of cellular DNA containing those integrated cosmid sequences to AT5BIVA cells produced cell clones with radioresistance similar to that of 1B3 cells, indicating that a gene within the cosmids complements the defect in the AT-D group [Kapp and
35 Painter 1989, id.].

Functional complementation of that same cell line (AT5BIVA) has been accomplished by microcell-mediated chromosome transfer from mouse-human hybrids [Lambert et al.

- 4 -

1991]. That study showed that the gene for AT-D was within a recombinant chromosome that contained a human chromosome 11q23 fragment that was telomeric to the AT-A and AT-C linkage region. However, the report did not state whether that mouse-human hybrid also contained the chromosome 11q23 region telomeric to THY1.

The inventors hereof cloned DNA from a fragment of an AT gene for complementation group D. That DNA was of value as a probe to find and clone an entire AT gene for complementation group D, that is, an ATDC gene, and to identify a region in 11q23 where the ATDC gene is located. Prior to the cloning described herein, basic research and clinical work on AT had moved very slowly as there had been no specific gene recognized for AT. Further, there had been no simple biochemical or laboratory test to identify AT patients and/or AT heterozygotes accurately or to classify patients into various complementation groups. The instant invention provides a clear direction for AT research and the means to identify mutations in ATDC genes. Identifying such mutations provides for methods to diagnose AT, preferably AT-D, and to detect AT heterozygotes, preferably AT-D heterozygotes. Detection of AT heterozygotes is important because they have been reported to have an increased risk of cancer in response to treatment with ionizing radiation [Swift et al., N. Engl. J. Med., 325: 1831-1836 (1991)].

SUMMARY OF THE INVENTION

Herein disclosed is an AT gene for complementation group D, the ATDC gene. Fragments of said gene are useful as nucleic acid probes and as polymerase chain reaction (PCR) primers. One embodiment of said ATDC gene comprises the human DNA in cosmid K1 and 4-1 which were deposited at the American Type Culture Collection [ATCC; Rockville, MD (USA)] on June 16, 1992, respectively under ATCC Nos. 75250 and 75251. This invention thus concerns, in one aspect, compositions comprising an isolated ATDC gene or one or more fragments thereof; the nucleotide sequence for said gene or fragment(s) thereof is or are substantially complementary to the

- 5 -

nucleotide sequence of the ATDC gene contained in cosmid K1 and/or cosmid 4-1, or to fragments of said nucleotide sequence contained within either or both of said cosmids.

Further disclosed is cDNA from said ATDC gene,
5 preferably a 3 kilobase (kb) cDNA and/or fragments thereof,
the entire nucleotide sequence for which is shown in Figures
4A-4C [SEQ. ID. NO.: 1]. Still further disclosed are AT
proteins and/or polypeptides translated from said cDNA,
preferably from said 3 kb cDNA. The amino acid sequence for a
10 preferred AT protein is shown in Figures 5A-5D [SEQ. ID. NO.:
3]. Portions of said amino acid sequence are included in this
invention as well as homologous variants of said amino acid
sequence and portions of said homologous variants. Within the
scope of this invention are AT proteins and/or polypeptides,
15 preferably ATDC proteins and/or polypeptides, produced
recombinantly, chemically and/or biologically, wherein said
proteins and/or polypeptides are in a substantially pure form.

In one facet, this invention concerns an isolated
nucleic acid sequence encoding a human AT protein/polypeptide
20 wherein the sequence is substantially complementary, that is,
hybridizable under standard stringency conditions, to a
nucleic acid sequence encoding the amino acid sequence
depicted in Figures 5A-5D, or at least a portion of said amino
acid sequence. Representative nucleic acid sequences encoding
25 such a human AT protein/polypeptide are those nucleic acid
sequences shown in Figures 4A-4C [SEQ. ID. NO.: 1] and in
Figures 5A-5D [SEQ. ID. NO.: 2].

Still further, this invention concerns recombinant
nucleic acid molecules comprising a genomic DNA sequence or a
30 cDNA sequence that encodes an AT protein or polypeptide,
wherein said DNA or cDNA sequence is operatively linked to an
expression control sequence in said nucleic acid molecule.
Exemplary embodiments of said recombinant nucleic acid
molecules include those wherein the genomic DNA sequence is
35 selected from the group consisting of the DNA sequence of the
ATDC gene contained in cosmid K1 and cosmid 4-1, DNA sequences
substantially complementary to the DNA sequence contained in
those cosmids, fragments of the DNA sequence contained in

- 6 -

those cosmids, and fragments of DNA sequences that are substantially complementary to the DNA sequence contained in said cosmids. Representative embodiments of said recombinant molecules also include those wherein the cDNA sequence is
5 selected from the group consisting of the 3 kilobase (kb) cDNA shown in Figures 4A-4C [SEQ. ID. NO.: 1], DNA sequences substantially complementary to said 3 kb cDNA, the cDNA sequence shown in Figures 5A-5D [SEQ. ID. NO.: 2], and DNA sequences substantially complementary to the cDNA sequence
10 shown in Figures 5A-5D.

Still further disclosed are polymerase chain reaction (PCR) primers for amplifying each of the 9 exons of said 3 kb cDNA. Said primers are delineated in Figure 7 [SEQ. ID. NOS.: 4-31], and the 14 PCR fragments amplified by said
15 primers are shown in Figures 8A-8N [SEQ ID. NOS.: 32-45].

Said ATDC gene, fragments thereof and/or the related cDNA are useful as follows: 1) to identify AT proteins and polypeptides as well as homologs or near-homologs thereto, preferably of the AT complementation group D; 2) to identify
20 mutations in ATDC genes leading to AT-D; 3) to identify various mRNAs transcribed from ATDC genes in various AT cell lines and from different human tissues; 4) to provide the means to construct probes to test for persons with AT and that are AT heterozygotes; 5) and to elucidate the mechanisms by
25 which AT mutations lead to radiosensitivity, immunodeficiencies and ataxia.

For example, fragments of the ATDC gene as contained in the deposited cosmids K1 and 4-1, or fragments thereof, or the 3 kb cDNA as shown in Figures 4A-4C and/or fragments
30 thereof, or the PCR primers for each of the nine exons, can be used as probes to identify and isolate ATDC genes from various AT patients. The nucleotide sequences of the ATDC gene from persons with AT and from those without AT can be compared, and the mutation(s) responsible for AT pleiotropic abnormalities
35 can be identified. Once the AT mutation(s) is or are identified, nucleic acid probes, preferably DNA probes, or PCR primers, can be constructed that span the mutated region(s). Such probes or PCR primers will provide the basis for the

development of tests, hybridization or PCR assays, to identify people with AT and heterozygote carriers of the mutated AT gene. PCR assays are preferred.

5 The sequence of the ATDC gene and/or fragments thereof and the resulting probes should allow for research to progress on the mechanism(s) underlying the pleiotropic abnormalities of AT and investigations into the reported high rates of some types of cancers in AT heterozygotes.

10 A representative method of detecting mutations in an ATDC gene according to this invention comprises the steps of:
amplifying one or more fragment(s) of said gene and of an ATDC gene known to be normal, by the polymerase chain reaction (PCR) with the same or substantially the same PCR primers; and

15 determining whether said ATDC gene contains any mutations by comparing the PCR products of the amplification of said gene with those from the amplification of the normal gene, and detecting differences between the PCR products associated with mutations. Preferred PCR primers for that
20 representative method are selected from the group consisting of the primers shown in Figure 7 [SEQ. ID. NOS.: 4-31] and nucleic acid sequences substantially complementary thereto. Differences detected by said method are, for example, differences in size and in base sequence. Preferred methods
25 of detecting mutations according to this invention include the use of a PCR-single-stranded conformation polymorphism assay or denaturing gradient electrophoretic assay.

BRIEF DESCRIPTION OF THE DRAWINGS

30 Figure 1A illustrates the structure of cosmid clones K1 and K41. The locations of the human DNA (thinner black lines) and integrated pCV108 cosmid sequences (undarkened bars) are shown. The neo gene, with an SV40 promoter and bifunctional termination sequences are positioned as shown.
35 The darkened bars indicate restriction fragments in the K1 and K41 cosmids that contain sequences complementary to cDNA from a HeLa cDNA library. The locations of the four EcoRI (E) fragments combined as probes for identification of cDNA clones

are shown, as are the locations of the restriction-enzymes sites for BglII (B), SmaI (Sm), SstI (Ss), XbaI (Xb), and XhoI (Xh).

Figure 1B illustrates the structure of the full-length 3.0 kilobase (kb) cDNA.

Figure 2A illustrates the structure of the 4-1 and 3-1 cosmids containing the 3' portion of the ATDC gene (the ataxia-telangiectasia complementation group D gene).

Figure 2B illustrates the structure of the intact ATDC gene, as reconstructed from the sequences in the K1 (left) and 4-1 (right) cosmids. The symbols for the DNA sequences and restriction sites are as defined in the description of Figures 1A-1B, supra.

Figure 3 provides a graph of survival as a function of X-ray dose for the following cell lines: AT5BIVA (indicated by blackened circle); 1B3 (indicated by unblackened circles); and C6 (indicated by blackened triangles).

Figures 4A-4C show the nucleotide sequence for the full-length 3 kb cDNA [SEQ. ID. NO.: 1], the restriction map for which is shown in Figure 1B.

Figures 5A-5D provide the translated open reading frame nucleotide sequence [SEQ. ID. No.: 2] for the 3 kb cDNA. [The amino acid sequence is SEQ. ID. NO. 3.]

Figure 6 schematically shows the nine exons of the 3 kb cDNA. Fourteen PCR fragments are delineated wherein the arrows indicate the positions of the primers. The primers are shown in Figure 7, and the sequences of the PCR fragments in Figures 8A-8N. The open box indicates the untranslated region whereas the hatched boxes indicate the region of the alpha helix.

Figure 7 provides the nucleotide sequences for the PCR primers [SEQ. ID. NOS.: 4-31] for the 3 kb cDNA, the placement of which is shown schematically in Figure 6 and specifically in Figures 8A-8N.

Figures 8A-8N provide the nucleotide sequences [SEQ. ID. NOS. 32-45] for the PCR fragments of the 3 kb cDNA schematically indicated in Figure 6 as well as the placement of the PCR primers listed in Figure 7.

Nucleotide and Amino Acid Sequence Symbols

The following symbols are used to represent nucleotides in the figures herein:

	<u>Base</u>	<u>Symbol</u>
5	adenine	A
	cytosine	C
	guanine	G
	thymine	T
	uracil	U

10

The symbol Y represents cytosine or thymine/uracil; whereas the symbol N represents adenine or cytosine or guanine or thymine/aracil according to the standard convention.

It is understood that because of the degeneracy of the genetic code, that is, that more than one codon will code for one amino acid [for example, the codons TTA, TTG, CTT, CTC, CTA and CTG each code for the amino acid leucine (leu)], that variations of the nucleotide sequence in, for example, Figures 4A-4C, wherein one codon is substituted for another, would produce a substantially equivalent protein or polypeptide according to this invention. All such variations in the nucleotide sequence of the 3 kb cDNA are included within the scope of this invention.

It is further understood that the nucleotide sequence herein described and shown in Figures 4A-4C represent only the precise structure of the cDNA nucleotide sequence isolated according to this invention. It is expected that slightly modified nucleotide sequences will be found or can be modified by techniques known in the art to code for substantially similar AT proteins and polypeptides, for example, those having similar epitopes, and such nucleotide sequences and proteins/polypeptides are considered to be equivalents for the purpose of this invention. DNA having equivalent codons is considered within the scope of the invention, as are synthetic DNA sequences that encode proteins/polypeptides homologous or substantially homologous to AT proteins/polypeptides, as well as those sequences but for the degeneracy of the genetic code would hybridize to said

- 10 -

3 kb cDNA nucleotide sequence. Modifications and variations of DNA sequences as indicated herein are considered to result in sequences that are substantially the same as the ATDC sequences and fragments thereof herein.

5 There are twenty main amino acids, each of which is specified by a different arrangement of three adjacent DNA nucleotides (triplet code or codon), and which are linked together in a specific order to form a characteristic protein. A three-letter convention is used herein to identify said
10 amino acids, as, for example, in Figures 5A-5D, as follows:

	<u>Amino acid name</u>	<u>Symbol</u>
	Alanine	Ala
	Arginine	Arg
15	Asparagine	Asn
	Aspartic Acid	Asp
	Cysteine	Cys
	Glutamic Acid	Glu
	Glutamine	Gln
20	Glycine	Gly
	Histidine	His
	Isoleucine	Ile
	Leucine	Leu
	Lysine	Lys
25	Methionine	Met
	Phenylalanine	Phe
	Proline	Pro
	Serine	Ser
	Threonine	Thr
30	Tryptophan	Trp
	Tyrosine	Tyr
	Valine	Val

Abbreviations

35 The following abbreviations are used herein.

AT - ataxia-telangiectasia
AT-A - ataxia-telangiectasia complementation group A
AT-B - ataxia-telangiectasia complementation group B

- 11 -

	AT-C	-	ataxia-telangiectasia complementation group C
	AT-D	-	ataxia-telangiectasia complementation group D
	ATCC	-	American Type Culture Collection
	<u>ATDC</u> gene	-	AT-D complementing gene
5	B	-	<u>Bgl</u> II
	DAPI	-	diamino-2-phenylindole
	E	-	<u>Eco</u> RI
	FITC	-	fluorescein isothiocyanate
	Gy	-	gray; unit of absorbed radiation according to
10			Système International d'Unités
	kb	-	kilobase
	M	-	molar
	min	-	minute
	ml	-	milliliter
15	PCR	-	polymerase chain reaction
	Sm	-	<u>Sma</u> I
	Ss	-	<u>Sst</u> I
	SV	-	simian virus
	μg	-	micrograms
20	Xb	-	<u>Xba</u> I
	Xh	-	<u>Xho</u> I

DETAILED DESCRIPTION

Herein disclosed is a new gene, an AT gene, more specifically the ATDC gene. That gene was isolated by complementation of ionizing radiation sensitivity in an AT cell line from complementation group D. Claimed herein is said ATDC gene and nucleotide sequences that are substantially complementary to said gene or fragments thereof. Strong evidence is provided herein to indicate that a defect in said gene is responsible for radiosensitivity in AT cells of complementation group D.

To isolate that gene, the radiosensitive AT cells of the AT5BIVA cell line were first transfected with a human cosmid library containing a selectable neo gene, followed by selection with both the antibiotic G418 and ionizing radiation to isolate cell clones containing human/cosmid DNA that complements the defective gene. A cell clone (1B3) that

- 12 -

demonstrates partial recovery of resistance to ionizing radiation was isolated. That clone was found to contain three integrated human/cosmid sequences.

The integrated human/cosmid sequences were rescued
5 by a process which first comprised the construction of a cosmid library from the 1B3 cell DNA and then the selection with kanamycin for clones containing the neo gene. The selected cosmid clones were then used as chromosome-specific painting probes [Pinkel et al., PNAS (USA), 85: 9138-9142
10 (Dec. 1988); Gray et al., European Patent Application 430,402 (published June 5, 1991); and Ward et al., WO 90/05789 (published May 31, 1990)] for in situ hybridization to chromosomes from normal human lymphocytes. Two related cosmid clones (K1 and K41) were found to contain human sequences that
15 originated from chromosomal region 11q23. As indicated in the Background above, the 11q23 region has been previously determined by genetic linkage analysis to be the location of the AT gene(s) from three separate complementation groups.

Restriction site mapping of the cosmid clones showed
20 that they contain over 36 kilobases (kb) of human DNA located to one side of an integrated cosmid sequence in cell clone 1B3. Southern blot analysis with total human DNA revealed that most of the 36 kb was single-copy DNA. A cDNA library was screened with parts of said single-copy DNA as probes, and
25 two clones that hybridized were isolated, indicating that the ATDC gene is transcribed in relative abundance. Transfection of one of the cosmids (K1) back in the radiosensitive parental AT cell line (AT5BIVA) resulted in the isolation of one cell clone that exhibited radio-resistance similar to that for 1B3.

30 Sequence analysis has shown that ATDC is a previously unidentified gene. Although the ATDC gene is a single-copy gene, five mRNAs have been detected in a range of from about 1.8 kilobases (kb) to about 5.7 kb. The variation in size of the mRNAs may result from differences at the 5'
35 end, indicating that the ATDC gene is a complex gene that can code for several different proteins. The ATDC gene may therefore have multiple functions, consistent with the pleiotropic abnormalities seen in individuals with AT.

Fine mapping of the location of the ATDC gene, as described below, by means of radiation hybrid cell lines, demonstrated that the ATDC gene is located just telomeric to the THY1 marker, that is, in the region that had been indicated by linkage analysis to contain the AT gene from complementation group D.

The ATDC gene and/or fragments thereof and/or cDNA therefrom may be used as probes to identify homologous genes from other AT complementation groups. Linkage analysis has demonstrated that the AT gene(s) for the various complementation groups are clustered together, and thus it is possible that they are distantly related. Chromosome 11-specific libraries, preferably chromosome 11q23-specific libraries, and still more preferably chromosome 11q23-specific cosmid libraries, can be screened to identify clones that cross hybridize with ATDC.

Further, mouse cDNA libraries can be screened with ATDC probes for a related mouse gene. Isolation of a mouse homolog would allow for the development of a transgenic mouse AT model.

The probes and assays enabled by this invention will be useful in proving that a defect in the ATDC gene is responsible for AT-D. Proof required would be (1) the demonstration of genetic rearrangements, such as, deletions, amplifications, translocations, inversions, point mutations among others, in the ATDC gene in AT-D cells, and/or (2) functional complementation of radiosensitivity in AT-D cells after transfection with an appropriate vector, preferably a cosmid vector containing the gene or a fragment thereof, or a cDNA therefrom in an expression vector. That there are no detectable rearrangements within the ATDC gene in the AT5BIVA cell line suggests a small rearrangement or point mutation, as observed in other human genetic diseases [Gibbs et al., Genomics, 7: 235-244 (1990); Groden et al., Cell, 66: 589-600 (1991)]. DNA from other AT cell lines from unknown complementation groups also showed no identifiable alterations in the ATDC gene. Preferably sequence analysis of the ATDC gene in cells obtained exclusively from AT-D individuals will

- 14 -

be used to determine whether mutations in the gene are associated with AT.

Assays to Detect Mutations in the ATDC Gene

5 An important utility of the instant invention is to detect the mutations that cause defect(s) in the ATDC gene resulting in the phenotypic symptoms of AT-D. To detect relatively large genetic rearrangements, hybridization tests, preferably Southern or Northern assays, can be used. To
10 detect relatively small genetic rearrangements, as for example, small deletions or amplifications, or point mutations, assays incorporating PCR are preferably used.

 Preferred methods of identifying mutations within the ATDC gene comprise assays wherein PCR is used. The
15 mechanics of PCR are described in Saiki et al., Science, 230: 1350 (1985) and U.S. Patent Nos. 4,800,159 (issued January 24, 1989), 4,683,195 and 4,683,202 (both of the latter issued July 28, 1987). Preferably, such a PCR assay amplifies and analyses cellular DNA, such that mutations within splice sites
20 as well as within coding regions can be detected when appropriate primers are used. However, mRNA could also be isolated, and cDNA prepared therefrom. Appropriate PCR primers could then be used to amplify the cDNA, and the cDNA PCR products from normal and AT cells could then be compared.

25 An exemplary PCR assay is that wherein the PCR primers listed in Figure 7 [SEQ. ID. NOS.: 4-31] are used. In such an assay, cellular DNA from people with AT and without AT is isolated and amplified with the PCR primers. The PCR products from the normal and AT cells are compared, preferably
30 initially upon a sizing gel. Changes in size indicative of certain genetic rearrangements are thus determined. If no changes are found in the size of the fragments, then further comparisons can be made to detect genetic rearrangements that would not be evident in such a size comparison, for example,
35 deletions of a few base pairs or point mutations. To effect such further comparisons, preferred methods include the use of a PCR-single-strand conformation polymorphism (PCR-SSCP) assay or a denaturing gradient gel electrophoretic assay. The PCR-

SSCP method is described in a review article by Kenshi Hayashi
entitled "PCR-SSCP: A Simple and Sensitive Method for
Detection of Mutations in the Genomic DNA" in PCR Methods and
Applications, 1: 34-38 (1991). Denaturing gradient gel
5 electrophoresis is described in Myers et al., "Detection and
Localization of Single Base Changes by Denaturing Gradient Gel
Electrophoresis," Methods in Enzymology, 155: 501-527 (1987).

Such assays, as hybridization and PCR assays,
preferably PCR assays including the PCR-SSCP or denaturing
10 gradient gel electrophoretic assays, as applied in the context
of this invention may be used to screen people for AT
heterozygosity, which can be a significant risk factor in
regard, for example, to ionizing radiation, and for prenatal
testing for and diagnosis of AT, among other uses.

15

AT Proteins and/or Polypeptides

The phrase "AT proteins and/or polypeptides" is
herein defined to mean proteins and/or polypeptides encoded by
an ATDC gene and/or fragments thereof. An exemplary and
20 preferred AT protein is that for which the amino acid sequence
is shown in Figures 5A-5D [SEQ. ID. NO.: 3].

A "polypeptide" is a chain of amino acids covalently
bound by peptide linkages and is herein considered to be
composed of 50 or less amino acids. A "protein" is herein
25 defined to be a polypeptide composed of more than 50 amino
acids.

It will be appreciated that the amino acid sequence
of AT proteins and polypeptides can be modified by genetic
techniques. One or more amino acids can be deleted or
30 substituted. Such amino acid changes may not cause any
measurable change in the biological activity of the protein or
polypeptide and result in proteins or polypeptides which are
within the scope of this invention.

The AT proteins and polypeptides of this invention
35 can be prepared in a variety of ways according to this
invention, for example, recombinantly, synthetically or
otherwise biologically, that is, by cleaving longer proteins

- 16 -

and polypeptides enzymatically and/or chemically. A preferred method to prepare AT proteins is by recombinant means.

A representative method to prepare the AT protein as shown in Figures 5A-5D or fragments thereof would be to insert the appropriate fragment of the 3 kb cDNA pictured in Figures 4A-4C into an appropriate expression vector. A wide variety of host-cloning vector combinations may be usefully employed in cloning the ATDC DNA isolated as described herein. For example, useful cloning vehicles may include chromosomal, nonchromosomal and synthetic DNA sequences such as various known bacterial plasmids such as pBR322, other E. coli plasmids and their derivatives and wider host range plasmids such as RP4, phage DNA such as the numerous derivatives of phage lambda, e.g., NB989 and vectors derived from combinations of plasmids and phage DNAs such as plasmids which have been modified to employ phage DNA expression control sequences.

Useful hosts may be eukaryotic or prokaryotic and include bacterial hosts such as E. coli and other bacterial strains, yeasts and other fungi, animal or plant hosts such as animal or plant cells in culture, insect cells and other hosts. Of course, not all hosts may be equally efficient. The particular selection of host-cloning vehicle combination may be made by those of skill in the art after due consideration of the principles set forth herein without departing from the scope of this invention.

The particular site chosen for insertion of the selected DNA fragment into the cloning vehicle to form a recombinant DNA molecule is determined by a variety of factors. These include size and structure of the protein or polypeptide to be expressed, susceptibility of the desired protein or polypeptide to endoenzymatic degradation by the host cell components and contamination by its proteins, expression characteristics such as the location of start and stop codons, and other factors recognized by those of skill in the art.

The recombinant nucleic acid molecule containing the ATDC gene, fragment thereof, or cDNA therefrom, may be

- 17 -

employed to transform a host so as to permit that host (transformant) to express the structural gene or fragment thereof and to produce the protein or polypeptide for which the hybrid DNA codes. The recombinant nucleic acid molecule
5 may also be employed to transform a host so as to permit that host on replication to produce additional recombinant nucleic acid molecules as a source of ATDC DNA and fragments thereof. The selection of an appropriate host for either of these uses is controlled by a number of factors recognized in the art.
10 These include, for example, compatibility with the chosen vector, toxicity of the co-products, ease of recovery of the desired protein or polypeptide, expression characteristics, biosafety and costs.

Where the host cell is a procaryote such as E. coli,
15 competent cells which are capable of DNA uptake are prepared from cells harvested after exponential growth phase and subsequently treated by the CaCl_2 method by well known procedures. Transformation can also be performed after forming a protoplast of the host cell.

20 Where the host used is an eucaryote, transfection method of DNA as calcium phosphate-precipitate, conventional mechanical procedures such as microinjection, insertion of a plasmid encapsulated in red blood cell hosts or in liposomes, treatment of cells with agents such as lysophosphatidyl-
25 choline or use of virus vectors, or the like may be used.

The level of production of protein or polypeptide is governed by two major factors: the number of copies of its gene or DNA sequence encoding for it within the cell and the efficiency with which these gene and sequence copies are
30 transcribed and translated. Efficiencies of transcription and translation (which together comprise expression) are in turn dependent upon nucleotide sequences, normally situated ahead of the desired coding sequence. These nucleotide sequences or expression control sequences define, inter alia, the location
35 at which RNA polymerase interacts to initiate transcription (the promoter sequence) and at which ribosomes bind and interact with the mRNA (the product of transcription) to initiate translation. Not all such expression control

- 18 -

sequences function with equal efficiency. It is thus of advantage to separate the specific coding sequences for the desired protein from their adjacent nucleotide sequences and fuse them instead to known expression control sequences so as to favor higher levels of expression. This having been achieved, the newly engineered DNA fragment may be inserted into a multicopy plasmid or a bacteriophage derivative in order to increase the number of gene or sequence copies within the cell and thereby further improve the yield of expressed protein.

Several expression control sequences may be employed. These include the operator, promoter and ribosome binding and interaction sequences (including sequences such as the Shine-Dalgarno sequences) of the lactose operon of E. coli ("the lac system"), the corresponding sequences of the tryptophan synthetase system of E. coli ("the trp system"), a fusion of the trp and lac promoter ("the tac system"), the major operator and promoter regions of phage lambda ($O_L P_L$ and $O_R P_R$), and the control region of the phage fd coat protein. DNA fragments containing these sequences are excised by cleavage with restriction enzymes from the DNA isolated from transducing phages that carry the lac or trp operons, or from the DNA of phage lambda or fd. These fragments are then manipulated in order to obtain a limited population of molecules such that the essential controlling sequences can be joined very close to, or in juxtaposition with, the initiation codon of the coding sequence.

The fusion product is then inserted into a cloning vehicle for transformation of the appropriate hosts and the level of antigen production is measured. Cells giving the most efficient expression may be thus selected. Alternatively, cloning vehicles carrying the lac, trp or lambda P_L control system attached to an initiation codon may be employed and fused to a fragment containing a sequence coding for an AT protein or polypeptide such that the gene or sequence is correctly translated from the initiation codon of the cloning vehicle.

The phrase "recombinant nucleic acid molecule" is herein defined to mean a hybrid nucleotide sequence comprising at least two nucleotide sequences, the first sequence not normally being found together in nature with the second.

5 The phrase "expression control sequence" is herein defined to mean a DNA sequence of nucleotides that controls and regulates expression of structural genes when operatively linked to those genes.

10 Synthetic and Biologic Production of AT Proteins and Polypeptides

AT proteins and polypeptides of this invention may be prepared not only by recombinant means but also by synthetic and by other biologic means. Exemplary of other
15 biologic means to prepare the desired polypeptide or protein is to subject to selective proteolysis a longer AT polypeptide or protein containing the desired amino acid sequence; for example, the longer polypeptide or protein can be split with chemical reagents or with enzymes. Synthetic formation of the
20 polypeptide or protein requires chemically synthesizing the desired chain of amino acids by methods well known in the art.

Chemical synthesis of a peptide is conventional in the art and can be accomplished, for example, by the Merrifield solid phase synthesis technique [Merrifield, J.,
25 Am. Chem. Soc., 85: 2149-2154 (1963); Kent et al., Synthetic Peptides in Biology and Medicine, 29 f.f. eds. Alitalo et al., (Elsevier Science Publishers 1985); and Haug, ABL, 40-47 (Jan/Feb. 1987)].

Techniques of chemical peptide synthesis include
30 using automatic peptide synthesizers employing commercially available protected amino acids, for example, Biosearch [San Rafael, CA (USA)] Models 9500 and 9600; Applied Biosystems, Inc. [Foster City, CA (USA)] Model 430; Milligen [a division of Millipore Corp.; Bedford, MA (USA)] Model 9050; and Du
35 Pont's RAMP (Rapid Automated Multiple Peptide Synthesis) [Du Pont Compass, Wilmington, DE (USA)].

The following materials and methods were used in the examples illustrating this invention.

Cell Lines

The simian virus (SV)40-transformed cell lines LM217 (normal) and AT2SF (unknown AT complementation group) were established by transfection of primary human fibroblasts with pSVori- plasmid [Murnane et al., Exp. Cell Res., 158: 119-126 (1985); Murnane et al., Mol. Cell Biol., 6: 549-558 (1986)].

The SV40-transformed fibroblast cell line AT3BISV (AT-A) was provided by A. M. R. Taylor (Birmingham, United Kingdom).

Cell line 1B3 was derived from the SV40-transformed fibroblast cell line AT5BIVA (GM5849) [Kapp and Painter 1989, supra], which was obtained from the NIGMS Human Genetic Mutant Cell Repository [Camden, NJ (USA)]. HeLa cells (Hela S3; ATCC CCL 2.2) were obtained from the American Type Culture Collection [ATCC; Rockville, MD (USA)].

Cosmid Library Construction

The cosmid library from 1B3 was constructed according to standard procedures [Sambrook et al., "Molecular Cloning: A Laboratory Manual", 2d ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York; 1989)]. DNA from 1B3 was partially digested with MboI, and fragments of about 30 kilobases (kb) to about 50 kb were ligated to BamHI-digested pWE16 cosmid DNA [Stratagene, La Jolla, CA (USA); Wahl et al., PNAS, 84: 2160 (1987)]. Clones containing the neo (neomycin resistance) gene in the integrated pCV108 cosmid [Lau and Kan, PNAS, 80: 5225 (1983)] DNA originally used for transfection [Kapp and Painter (1989)] were selected by growth of the XL1-Blue bacteria (Stratagene) containing the cosmid library on agar plates with 50 µg kanamycin/ml. Restriction-enzyme mapping was accomplished by digestion with various combinations of enzymes and was confirmed by further mapping of the individual EcoRI fragments cloned into a Bluescript plasmid (Stratagene).

Cosmid and cDNA Library Screening

A commercially available HeLa cell cDNA library (Stratagene) within the bacteriophage vector ZAPII was

screened according to a method described elsewhere [Murnane, Mol. Cell Biol., 6: 549-558 (1986)] with a probe that consisted of cosmid fragments containing only nonrepetitive sequences. Positive cDNA clones were then rescued from ZAPII
5 by in vivo excision of the Bluescript plasmid by using a protocol provided with the vector (Stratagene). A chromosome 11-specific cosmid library obtained from L. Deaven [Los Alamos National Laboratory; Los Alamos, NM (USA)] was screened by conventional methods (Sambrook et al. 1989, supra) with the
10 use of a 3.0 kb cDNA as a probe (Figures 1B, 4A-4C and 8A-8N).

Southern and RNA Blot Analysis

Preparation of high-molecular-weight cellular DNA and Southern blot analysis were as described elsewhere
15 (Murnane 1986, supra). mRNA was prepared with the use of an mRNA isolation kit [Fast Track^R; Invitrogen; San Diego, CA (USA)]. RNA agarose gel electrophoresis and RNA blot analysis were carried out according to standard procedures (Sambrook et al. 1989, supra).

20

Functional Complementation

DNA to be tested for complementation was introduced into AT5BIVA cells by calcium phosphate-mediated transfection (Murnane et al. 1985, supra). Before transfection, the cosmid
25 clones were linearized with NotI, which cuts between the vector and human sequences. Cell clones containing the integrated DNA were selected by incubation with 400 µg G418/ml. X-ray survival was determined according to a method described elsewhere (Kapp and Painter 1989, supra) with or
30 without feeder layers that consisted of HeLa cells exposed to 60 Gy (gray) of X-rays.

In Situ Hybridization

Metaphase spreads were prepared according to the
35 method described by Yunis in Science, 191: 1268-1270 (1976). Slides were treated with 20 µg pepsin/ml (in 0.1 M HCl) for 10 min at 37°C before in situ hybridization. Hybridization conditions and staining procedures were modified from those of

- 22 -

Pinkel et al. [PNAS (USA), 83: 2934-2938 (1986)] and Trask et al. [Am. J. Hum. Genet., 48: 1-15 (1991)]. The in situ hybridization probe strategy specifically comprised the use of a dioxigenin-labeled K1 or K41 cosmid probe that was detected with FITC-conjugated antibodies, and a biotin-labeled chromosome 11-specific alpha-satellite DNA probe that was detected with Texas Red-conjugated avidin. The chromosome 11-specific alpha-satellite DNA probe was used as a marker in locating the origin of the cosmid probes to human chromosomal region 11q23. Chromosomes were counterstained with 4',6-diamino-2-phenylindole (DAPI).

Radiation Hybrid Mapping

The chromosomal map location of the ATDC gene was determined by a radiation hybrid mapping technique [Richard et al., Am. J. Hum. Genet., 49: 1189-1196 (1991)]. The order and distances between ATDC and the markers stromelysin 1 (STMY1), CJ52.193 (D11S384), CJ52.77 (D11S424), CD3D, apolipoprotein (APO), THY1, D11S528, or ETS1 were established by a statistical analysis of the cosegregation of markers in 100 radiation hybrids, and the method of moments was used to determine the frequency of breakage between markers [Cox et al., Science, 250: 245-250 (1990)]. Because a single retention frequency was used in this analysis, and since each hybrid was scored for almost every marker, the likelihood for any marker order was estimated as the sum of the pairwise lod scores between markers. The map with the highest likelihood was defined as that map in which the sum of lod scores between adjacent loci was maximized. The polymerase chain reaction (PCR) with primers specific for the 3' end of the 3.0 kb cDNA [primers 9D-5 and 9D-3 (SEQ. ID. NOS.: 30 and 31) shown in Figure 7] was used to determine which of the radiation hybrids contained the ATDC gene. The conditions for PCR were the same as those used for mapping other markers on human chromosome 11q (Richard et al. 1991, supra).

The following examples are for purposes of illustration only and are not meant to limit the invention in any way.

Example 1

Isolation of Integrated Cosmid DNA

The integrated cosmid sequences from cell line 1B3 were isolated by construction of a cosmid library from 1B3 DNA, and then by using kanamycin to select bacteria that had cosmid clones containing the integrated neo gene. Cosmid clones isolated by this method were screened by in situ hybridization to identify cosmid clones containing human DNA from the chromosomal region 11q23, previously demonstrated to be the location of the gene(s) associated with three AT complementation groups. Two of the cosmid clones (K1 and K41) hybridized to chromosomal region 11q23. As indicated above, a dioxigenin-labeled K1 or K41 cosmid probe was localized to 11q23 in a chromosome-specific painting experiment wherein a chromosome 11-specific alpha-satellite DNA probe was used concurrently to identify the centromeric chromosome 11 region.

The rescue from the integration site in 1B3 of DNA that originated from the chromosomal region 11q23 is further strong evidence that a gene within the transfected DNA complements radiosensitivity in cells from AT-D. According to restriction-enzyme mapping, K1 and K41 contained overlapping sequences, consisting of both the integrated pCV108 cosmid and adjacent human DNA (Figure 1A).

To determine whether the cloned human DNA contained any expressed cellular genes, fragments of cosmid K1 were used as probes to screen a HeLa cell cDNA library. Four adjacent EcoRI fragments that contained only nonrepetitive cellular DNA (Figure 1A) were combined for this purpose. After hybridization to 1.2×10^6 cDNA clones, 40 positive plaques were identified. Restriction-enzyme mapping of nine of these cDNA clones demonstrated that they were all related and that the five largest (3.0 kb) were nearly identical (Figure 1B). Partial sequence analysis of the ends of the 3.0-kb cDNAs has identified the 3' end by the presence of a tract of poly-A and suggests that they are full length because the 5' ends show two stop codons before the first methionine codon in the only complete open reading frame. [Full nucleotide sequence of 3 kb cDNA is shown in Figures 4A-4C.] That gene contained

within those cosmids is herein termed the "AT-D complementing (ATDC) gene."

The approximate position of the sequences complementary to the cDNA within the K1 cosmid was determined by Southern blot analysis. Southern blots of K1 cosmid DNA digested with various restriction enzymes were hybridized with either the full-length 3.0-kb cDNA or a fragment from the 5' end. By the location of the labeled restriction fragments, two separate regions of hybridization were identified (Figure 1A). Hybridization with the fragment from the 5' end of the cDNA indicated that the direction of transcription in the cosmid was from left to right (Figure 1A). Consistent with those results are the similarity of the restriction sites (XhoI, SstI, and SmaI) between the cDNA and cosmid clones (Figure 1), as well as partial sequence analysis, which also demonstrated that the region of hybridization to the left (Figure 1A) coded for the 5' end of the transcript.

Not all of the sequences found in the cDNA were contained within the K1 and K41 cosmids. This was evident from the absence of additional hybridization regions containing SstI and EcoRI restriction sites corresponding to those found within the cDNA and indicates that nearly half of the coding sequences within the cDNA were absent from these cosmids. The fact that the transfected pCV108 sequences were located adjacent to the ATDC gene in both the K1 and K41 cosmid clones shows that the truncated gene is the form integrated into cell line 1B3. The orientation of the integrated sequences placed the SV40 bidirectional transcriptional termination sequences, which are located adjacent to the neo gene in the pCV108 cosmid, down-stream from the truncated gene (Figure 1A). The form of the ATDC gene integrated into cell line 1B3 is therefore a functional transcriptional unit despite the absence of the 3' end.

The presence of only a portion of the ATDC gene in 1B3 can explain several properties of 1B3, namely: (a) radiosensitivity in 1B3 did not completely return to the level seen in normal cells [Kapp and Painter 1989, supra]; (b) 1B3 amplified the integrated DNA, whereas other independently

derived, related clones that did not amplify the integrated cosmid sequences failed to maintain radioresistance in culture [Kapp and Painter 1989]; and (c) 1B3 cells continued to show radioresistant DNA synthesis similar to that of the parental
5 AT5BIVA cell line [Kapp and Painter 1989].

Example 2

Isolation of the 3' End of the ATDC Gene and Composite Mapping of the ATDC Gene

10 To obtain the missing portion of the ATDC gene, the cDNA was used as a probe to screen a human genomic chromosome 11 library. Two cosmid clones (3-1 and 4-1) were identified. Restriction mapping of these clones demonstrated that they
15 the 3.0-kb cDNA (Figures 1B and 4A-4C) as a probe, Southern blot analysis revealed three separate regions of hybridization within the cosmid DNA (Figure 2A). One of these regions, which was contained within cosmid clone 4-1 but not within 3-1, overlapped with cosmid clones K1 and K41 (Figure 2A). That
20 was determined by sequence analysis with primers derived from the cDNA sequence. Beyond the overlapping region (compare Figures 1A and 2A), the differences between K1 and 4-1 are apparently due to the rearrangements that occurred in K1 during transfection and that resulted in the loss of the 3'
25 end of the gene. Cosmid clone 4-1 therefore contains the ATDC gene 3' portion that was missing from clones K1 and K41. A composite map derived from cosmid clones K1 and 4-1 indicates that the missing 3' portion of the ATDC gene is more than 30 kb in length (Figure 2B).

30 Southern blot analysis of human genomic DNA from normal cells with the 3.0-kb cDNA as a probe was consistent with the structure of the gene predicted from restriction-enzyme mapping analysis of the cosmid clones (Figure 2B). Bands detected after digestion of human DNA with BglII (1.8,
35 4.5, 8.0, and 15 kb), EcoRI (4.0, 6.0, and 20kb), and XbaI (9.0, 12, and 23kb) all corresponded to the expected fragments. [Lambda HindIII restriction fragments were used as markers.] The 1.8-kb BglII fragment contains only a small

- 26 -

number of complementary sequences and is difficult to detect. The absence of additional bands demonstrates that no other genes complementary to the ATDC gene are present in the human genome. Southern blot analysis of DNA isolated from the parental SV40-transformed AT fibroblast cell line, AT5BIVA, as well as two other SV40-transformed AT fibroblast cell lines (AT3BISV, AT-A; AT2SF, unknown complementation group) gave results identical to those from the normal cell DNA.

Example 3

Characterization of mRNA Transcripts

Although ATDC is a single-copy gene, it produces several mRNAs of various sizes, apparently because of alternate processing of RNA transcripts. RNA blot analysis of poly-A-selected mRNA isolated from an SV40-transformed fibroblast cell line (LM217) demonstrated two mRNA transcripts (5.7 and 4.7 kb) that hybridized to the 3.0-kb cDNA probe. In contrast, mRNA isolated from HeLa cells upon RNA blot analysis lacked the largest mRNA transcript (which was seen more clearly on shorter exposures), and instead contained at least two others (3.0 and 1.8 kb). In HeLa cells, the 3.0-kb mRNA transcript was the most abundant, consistent with the fact that the cDNA for this mRNA was the only one identified in the HeLa cell library (Figure 1B).

Two SV40-transformed AT fibroblast cell lines (AT3BISV and AT5BIVA) were shown upon RNA blot analysis to contain various amounts of the mRNA transcripts seen in the LM217 and HeLa cells. In contrast, the 1B3 cell line appeared identical to the SV40-transformed normal line LM217, in size of mRNAs. The significance of the differences between the mRNA observed in the AT3BISV and AT5BIVA cell lines and that seen in LM217 or 1B3 is not known. No additional mRNA transcript from the truncated ATDC gene was evident in 1B3, indicating that it is either below the level of detection or is similar in size to one of the mRNAs transcribed from the endogenous gene.

The production of multiple mRNAs has been observed with other mammalian genes and can result from multiple

- 27 -

promoters, alternate splicing, and/or alternate poly-A addition sites [Leff et al., Ann. Rev. Biochem., 55: 1091-1117 (1986)]. The presence of different mRNAs in epithelial (HeLa) and fibroblast (LM217, 1B3, AT5BIVA, and AT3BISV) cell lines, as shown by RNA analysis of poly-A-selected mRNA, could indicate tissue-specific processing of ATDC mRNA [Leff et al., 1986, id.] which would be consistent with the pleiotropic characteristics of AT. Because the 3.0-kb mRNA that produced the cDNA that was isolated from HeLa cells is not found in the LM217 cell line, functional complementation of radioresistance may require the cDNA from the 4.7-kb mRNA, which is present in both HeLa and LM217.

Example 4

Functional Complementation

After transfection of AT5BIVA cells with the K1 cosmid, 50 G418-resistant clones were tested for sensitivity to X-rays. Three clones showed increased radioresistance; one of them (C6) had about the same radioresistance as did the 1B3 cell line (Figure 3).

Thus, these initial functional complementation studies achieved a partial restoration of radioresistance in the AT5BIVA cell line as shown in Figure 3. The absence of full complementation with the K1 cosmid could be due to several factors. The large size of the ATDC gene may make complementation difficult because of the problems associated with integrating large intact DNA fragments into human cells [Colbere-Garapin et al., Gene, 50: 279-288 (1986); Hoeijmakers et al., Exp. Cell Res., 169: 111-119 (1987); Mayne et al., Gene, 66: 65-76 (1988) (erratum: Gene, 83: 395 (1989))]. Furthermore, because the ATDC gene produces mRNAs larger than the 3.0-kb mRNA characterized herein, it is unclear whether the K1 cosmid contains all of the coding sequences transfected into 1B3.

- 28 -

Example 5Fine Mapping of the ChromosomalLocation of the ATDC Gene

Radiation hybrids that contained mapped fragments of human chromosome 11 (Richard et al. 1991, supra) were used to determine the location of the ATDC gene. The position of the ATDC gene relative to known markers within 11q23 was established by using PCR to identify which of the radiation hybrids contained the gene. Statistical analysis of 100 radiation hybrid cell lines showed that the ATDC gene is closely linked to THY1 and D11S528, with lod scores of 12.2 and 17, respectively (Table 1). The order 11cen-APO-CD3D-THY1-ATDC-11qtel was 1,000 times more likely than the order 11cen-APO-CD3D-ATDC-THY1-11qtel which places the ATDC gene centromeric to THY1.

Thus, the ATDC gene lies outside the linkage region predicted to contain the AT gene(s) for complementation groups A and C, indicating a separate locus for the complementation group D gene. The results described herein clearly demonstrate that the ATDC gene is not linked to the centromeric (STMY1) and telomeric (D11S424) flanking markers for the AT-A and AT-C gene(s). However, the evidence that the ATDC gene isolated from 1B3 is telomeric to THY1 is consistent with the evidence for an additional AT gene within that region. It appears from the evidence that several genes within the 11q23 region are associated with AT; the lack of hybridization of ATDC to other human DNA sequences suggests that they are not closely related.

Table 1
Radiation Hybrid Two-Point Mapping Analysis
of ATDC and Various Chromosome 11 Loci

	Marker		CR ₉₀₀₀ ^a	LOD ^b	No. of Clones Observed ^c						Total
	A	B			+	+	+	-	-	-	
5	ATDC	STMY15	7	13	15	64			99
10	ATDC	D11S384 .		1.4	9	11	14	65			99
	ATDC	D11S424 .		1.0	8	12	13	66			99
	ATDC	APO 81		3.1	10	10	8	71			99
	ATDC	CD3D 37		7.7	13	7	3	76			99
	ATDC	THY1 17		12.2	15	5	0	79			99
15	ATDC	D11S528 . 6		17.0	18	2	0	79			99
	ATDC	ETS1 28		2.0	9	11	10	69			99
	APO	CD3D 50		9.5	14	5	3	78			100
	APO	THY1 17		5.3	11	8	5	76			100
	CD3D	THY1 17		11.7	14	3	2	81			100

^a Represents distances (in centiRays), which are shown only for linked markers (LOD > 3).

^b Likelihood that two markers are linked.

^c Number of hybrids that retain both markers A and B (+ +), marker A but not marker B (+ -), marker B but not marker A (- +), or neither marker A nor B (- -); total number of hybrids analyzed is also shown.

The materials listed below were deposited with the American Type Culture Collection (ATCC) in Rockville, MD (USA). The deposits were made under the provisions of the Budapest Treaty on the International Recognition of Deposited Microorganisms for the Purposes of Patent Procedure and Regulations thereunder (Budapest Treaty). Maintenance of a viable culture is assured for thirty years from the date of deposit. The organism will be made available by the ATCC

under the terms of the Budapest Treaty, and subject to an agreement between the Applicants and the ATCC which assures unrestricted availability upon issuance of the pertinent U.S. Patent. Availability of the deposited strain is not to be
 5 construed as a license to practice the invention in contravention of the rights granted under the authority of any Government in accordance with its patent laws.

	<u>Cosmid</u>	<u>Deposit Date</u>	<u>ATCC #</u>
10	K1	June 16, 1992	75250
	4-1	June 16, 1992	75251

The description of the foregoing embodiments of the invention have been presented for purposes of illustration and description. They are not intended to be exhaustive or to
 15 limit the invention to the precise form disclosed, and obviously many modifications and variations are possible in light of the above teachings. The embodiments were chosen and described in order to explain the principles of the invention and its practical application to enable thereby others skilled
 20 in the art to utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto.

All references cited herein are hereby incorporated
 25 by reference.

CLAIMS

We claim:

1. A composition comprising an isolated gene for ataxia-telangiectasia complementation group D (ATDC) or a fragment thereof, wherein the nucleotide sequence for said gene or for said fragment is the nucleotide sequence of the ATDC gene or a fragment of that ATDC gene that is contained in cosmid K1 and cosmid 4-1, deposited at the American Type Culture Collection (ATCC) under ATCC Nos. 75250 and 75251, respectively; or is substantially complementary to that ATDC gene sequence contained in said cosmids or to fragments of said ATDC gene sequence.

2. A composition comprising cosmid K1 and/or cosmid 4-1 deposited at the American Type Culture Collection (ATCC) under ATCC Nos. 75250 and 75251, respectively.

3. A composition comprising an isolated nucleic acid sequence encoding a human AT protein/polypeptide selected from the group consisting of: the cDNA sequence shown in Figures 4A-4C [SEQ. ID. NO.: 1], sequences substantially complementary to said cDNA sequence [SEQ. ID. NO.: 1], fragments of said cDNA sequence [SEQ. ID. NO.: 1], and fragments that are substantially complementary to fragments of said cDNA sequence [SEQ. ID. NO.: 1].

4. A composition according to Claim 3 wherein said one or more fragments of said cDNA sequence [SEQ. ID. NO.: 1] are the fragments shown in Figures 8A-8N [SEQ. ID. NOS.: 32-45], and the fragments that are substantially complementary to fragments of said cDNA sequence [SEQ. ID. NO.: 1] are nucleic acid sequences that are substantially complementary to the fragments shown in Figures 8A-8N [SEQ. ID. NOS.: 32-45].

- 32 -

5. A composition comprising polymerase chain reaction (PCR) primers selected from the group consisting of the primers shown in Figure 7 [SEQ. ID. NOS.: 4-31] and nucleic acid sequences substantially complementary thereto.

6. A composition comprising a protein or a polypeptide encoded by an ATDC gene or fragment thereof according to Claim 1.

7. A composition comprising an ataxia-telangiectasia (AT) protein or polypeptide having an amino acid sequence selected from the group consisting of: the amino acid sequence shown in Figures 5A-5D [SEQ. ID. NO.: 3]; a portion of the amino acid sequence shown in Figures 5A-5D; amino acid sequences which are a homologous variants of the amino acid sequence shown in Figures 5A-5D; and portions of amino acid sequences which are homologous variants of the amino acid sequence shown in Figures 5A-5D.

8. A recombinant nucleic acid molecule comprising a genomic DNA or cDNA sequence that encodes an AT protein or polypeptide according to Claim 7, wherein said genomic DNA or cDNA sequence is operatively linked to an expression control sequence in said nucleic acid molecule.

9. The recombinant nucleic acid molecule according to Claim 8 wherein said genomic DNA sequence is selected from the group consisting of: the DNA sequence of the ATDC gene contained in cosmid K1 and cosmid 4-1, deposited at the American Type Culture Collection (ATCC) under ATCC Nos. 75250 and 75251, respectively; DNA sequences substantially complementary to said DNA sequence contained in said cosmids; fragments of said DNA sequence contained in said cosmids; and DNA sequences substantially complementary to fragments of the DNA sequence contained in said cosmids.

10. The recombinant nucleic acid molecule according to Claim 10 wherein said cDNA sequence is selected from the

group consisting of: the 3 kilobase (kb) cDNA shown in
Figures 4A-4C [SEQ. ID. NO.: 1]; DNA sequences substantially
complementary to said 3 kb cDNA; the cDNA sequence shown in
Figures 5A-5D [SEQ. ID. NO.: 2]; and DNA sequences
5 substantially complementary to said cDNA sequence shown in
Figures 5A-5D.

11. A method of detecting mutations in an ATDC gene
comprising the steps of:
10 amplifying one or more fragment(s) of said gene and
of an ATDC gene known to be normal, by the polymerase chain
reaction (PCR) with the same or substantially the same PCR
primers; and

determining whether said ATDC gene contains any
15 mutations by comparing the PCR products of the amplification
of said gene with those from the amplification of the normal
gene, and detecting differences between the PCR products
associated with mutations.

12. A method according to Claim 11 wherein said PCR
20 products are compared for differences in size.

13. A method according to Claim 11 comprising the
use of a PCR-single-strand conformation polymorphism assay or
25 a denaturing gradient gel electrophoretic assay to determine
whether said ATDC gene contains any mutations.

14. A method according to Claim 11 wherein said PCR
primers are selected from the group consisting of the primers
30 shown in Figure 7 [SEQ. ID. NOS.: 4-31] and nucleic acid
sequences substantially complementary thereto.

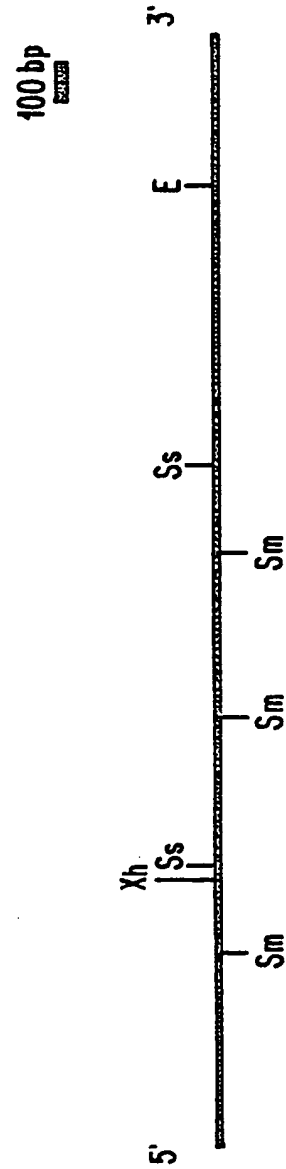
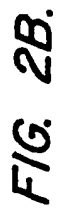


FIG. 1B.



3/21

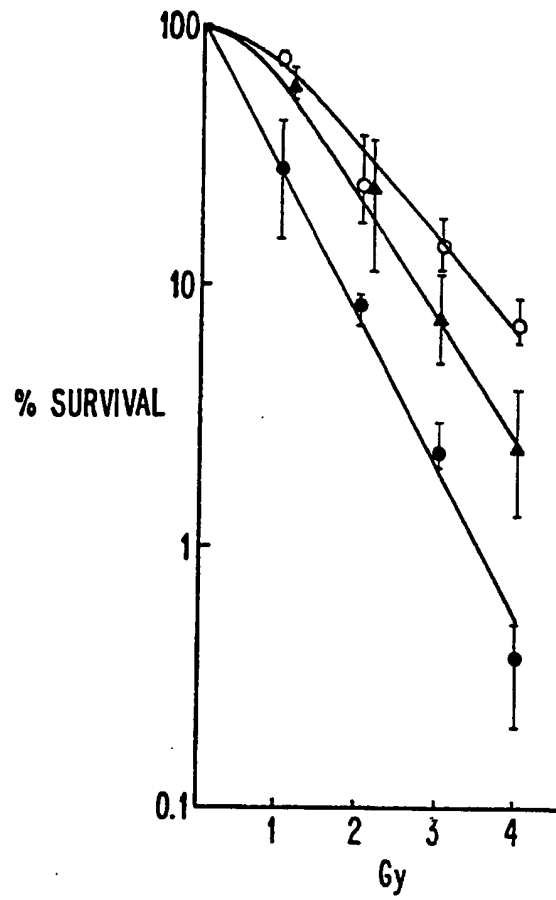


FIG. 3.

4/21

1 CTCCTCACAG GTGTGTCTCT AGTCCTCGTG GTTGCCTGCC CCACTCCCTG

51 CCGAGACGCC TGCCAGAAAG GTCACCTATC CTGAACCCCA GCAAGCCTGA

101 AACAGCTCAG CCAAGCACCC TGGGATGGAA GCTGCAGATG CCTCCAGGAG

151 CAACGGGTCC AGCCAGAAAG CCAGGGATGC CCGGAGCCCG TCGGGCCCCA

201 GTGGCAGCCT GGAGAATGGC ACCAAGGCTG ACGGCAAGGA TGCCAAGACC

251 ACCAACGGGG ACGGCGGGGA GGCAGCTGAG GGCAGAGGCC TGGGCAGCGC

301 CCTGAAGCCA GGGGAAGGTA GGAGCGCCCT GTTCGGGGGC AATGAGTGGC

351 GGGGACCCAT CATCCAGTTT GTCGAGTCCG GGGACGACAA GAACTCCAAC

401 TACTTCAGCA TGGACTCTAT GGAAGGCAAG AGGTCGCCGT ACGCAGGGCT

451 CCAGCTGGGG GCTGCCAAGA AGCCACCCGT TACCTTTGCC GAAAAGGGCG

501 ACGTGGGCAA GTCCATTTTC TCGGAGTCCC GGAAGCCAC GGTGTCCATC

551 ATGGAGCCCG GGGAGACCCG GCGGAACAGC TACCCCGGGG CCGACACGGG

601 CCTTTTTTCA CGGTCCAAGT CCGGCTCCGA GGAGGTGCTG TCGGACTCCT

651 GCATCGGCAA CAAGCAGAAG GCGGTCAAGT CCTGCCTGGT GTGCCAGGCC

701 TCCTTCTGG AGCTGCATCT CAAGCCGCAC CTGGAAGGGCG CGGCCTTCCG

751 AGACCACCAG CTGCTCGAGC CCATCCGGGA CTTTGAGGCC CGCAAGTGTC

801 CCGTGCATGG CAAGACGATG GAGCTCTTCT GCCAGACCGA CCAGACCTGC

851 ATCTGCTACC TTTGCATGTT CCAGGAGCAC AAGAATCATA GCACCGTGAC

901 AGTGGAGGAG GCCAAGGCCG AGAAGGAGAC GGAGCTGTCA CTGCAAAAGG

951 AGCAGCTGCA GCTCAAGATC ATTGAGATTG AGGATGAAGC TGAGAAGTGG

1001 CAGAAGGAGA AGGACCGCAT CAAGAGCTTC ACCACCAATG AGAAGGCCAT

1051 CCTGGAGCAG AACTTCCGGG ACCTGGTGGC GGACCTGGAG AAGCAAAAGG

1101 AGGAAGTGAG GGCTGGGCTG GAGCAGCGGG AGCAGGATGC TGTGGACCAA

1151 GTGAAGGTGA TCATGGATGC TCTGGATGAG AGAGCCAAGG TGCTGCATGA

1201 GGACAAGCAG ACCCGGGAGC AGCTGCATAG CATCAGCGAC TCTGTGTTGT

1251 TTCTGCAGGA ATTTGGTGCA TTGATGAGCA ATTACTCTCT CCCCCACCC

1301 CTGCCACCT ATCATGTCCT GCTGGAGGGG GAGGGCCTGG GACAGTCACT

FIG. 4A.

5/21

1351 AGGCAACTTC AAGGACGACC TGCTCAATGT ATGCATGCGC CACGTTGAGA
 1401 AGATGTGCAA GGCGGACCTG AGCCGTAAC TATTGAGAG GAACCACATG
 1451 GAGAACG|GTG ^{EXON 4} ^{EXON 5} GTGACCATCG CTATGTGAAC AACTACACGA ACAGCTTCGG
 1501 GGGTGAGTGG AGTGACCCGG ACACCATGAA GAGATACTCC ATGTACCTGA
 1551 CACCCAAAG|G ^{EXON 5} ^{EXON 6} TGGGGTCCGG ACATCATACC AGCCCTCGTC TCCTGGCCGGC
 1601 TTCACCAAGG AGACCACCCA GAAGAATTTC AACAATCTCT ATGGCACCAG
 1651 ^{EXON 6} ^{EXON 7} AG|GTAAC TAC ACCT|CCGGG ^{SmaI} TCTGGGAGTA CTCCTCCAGC ATTCAGAACT
 1701 CTGACAATGA CCTGCCCGTC GTCCAAGGCA GCTCCTCCTT CTCCTGAAA
 1751 ^{EXON 7} ^{EXON 8} G|GCTATCCCT CCCTCATGCG GAGCCAAAGC CCAAGGCCC AGCCCCAGAC
 1801 TTGAAATCT GGCAAGCAGA CTATGCT|GTC ^{EXON 8} ^{EXON 9} TCACTACCGG CCATTCTAGG
 1851 TCAACAAAGG CAACGGGATT GGGTCCAACG AAGCCCCATG ^{STOP} ^{SstI} AGCTCCTGGC
 1901 GGAAGGAACG AGGCGCCACA CCCCTGCTCT TCCTCCTGAC CCTGCTGCTC
 1951 TTGCCTTCTA AGCTACTGTG CTTGTCTGGG TGGGAGGGAG CCTGGTCTTG
 2001 CACCTGCCCT ^{Pst I} CTGCAGCCCT CTGCCAGCCT CTTGGGGGCA GTTCCGGCCT
 2051 CTCCGACTTC CCCACTGGCC ACACTCCATT CAGACTCCTT TCCTGCCTTG
 2101 TGACCTCAGA TGGTCACCAT CATTCTGTG CTCAGAGGCC AACCCATCAC
 2151 AGGGGTGAGA TAGGTTGGGG CCTGCCCTAA CCGGCCAGCC TCCTCCTCTC
 2201 GGGCTGGATC TGGGGGCTAG CAGTGAGTAC CCGCATGGTA TCAGCCTGCC
 2251 TCTCCGGCCC AGGCCCTGCT GTCTCCAGGC CTATAGACGT TTCTCTCAA
 2301 GGCCCTATCC CCAATGTTG TCAGCAGATG CCTGGACAGC ACAGCCACCC
 2351 ATCTCCCAT TACATGGCCC ACCTCCTGCT TCCAGAGGA CTGGCCCTAC
 2401 GTGCTCTCTC TCGTCCTAGC TATCAATGCC CAGCATGGCA ^{PstI} GAACCTGCAG
 2451 TGGCCAAGGG ^{PstI} CTGCAGATGG AAACCTCTCA GTGTCTTGAC ATCACCCTAC
 2501 CCAAGCGGTG GGTCTCCACC ACAGCCACTT TGAGTCTGTG GTCCCTGGAG
 2551 GGTGGCTTCT CCTGACTGGC AGGATGACCT TAGCCAAGAT ATTCCTCTGT
 2601 TCCCTCTGCT GAGATAAGA ^{EcoRI} ATTCCTTAA CATGATATAA TCCACCCATG
 2651 CAAATAGCTA CTGGCCAGC TACCATTAG CATTTGCCTA CAGAATTTCA

FIG. 4B.

2701 TTCAGTCTAC ACTTTGGCAT TCTCTCTGGC GATGGAGTGT GGCTGGGCTG
2751 ACCGCAAAAG GTGCCTTACA CACTGCCCCC ACCCTCAGCC GTTGCCCCAT
2801 CAGAGGCTGC CTCCTCCTTC TGATTACCCC CCATGTTGCA TATCAGGGTG
2851 CTCAAGGATT GGAGAGGAGA CAAAACCAGG AGCAGCACAG TGGGGACATC
2901 TCCCGTCTCA ACAGCCCCAG GCCTATGGGG GCTCTGGAAG GATGGGCCAG
2951 CTTGCAGGGG TTGGGGAGGG AGACATCCAG CTTGGGCTTT CCCCTTTGGA
POLY A ADDITION SEQUENCE
3001 ATAAACCATT GGTCTGTC - POLY A

FIG. 4C.

7/21

1 21 41
 ATG GAA GCT GCA GAT GCC TCC AGG AGC AAC GGG TCG AGC CCA GAA
 1: met glu ala ala asp ala ser arg ser asn gly ser ser pro glu

61 81
 GCC AGG GAT GCC CGG AGC CCG TCG GGC CCC AGT GGC AGC CTG GAG
 1: ala arg asp ala arg ser pro ser gly pro ser gly ser leu glu

101 121
 AAT GGC ACC AAG GCT GAC GGC AAG GAT GCC AAG ACC ACC AAC GGG
 1: asn gly thr lys ala asp gly lys asp ala lys thr thr asn gly

141 161
 CAC GGC GGG GAG GCA GCT GAG GGC AAG AGC CTG GGC AGC GCC CTG
 1: his gly gly glu ala ala glu gly lys ser leu gly ser ala leu

181 201 221
 AAG CCA GGG GAA GGT AGG AGC GCC CTG TTC GCG GGC AAT GAG TGG
 1: lys pro gly glu gly arg ser ala leu phe ala gly asn glu trp

241 261
 GGG CGA CCC ATC ATC CAG TTT GTC GAG TCC GGG GAC GAC AAG AAC
 1: arg arg pro ile ile gln phe val glu ser gly asp asp lys asn

281 301
 TCC AAC TAC TTC AGC ATG GAC TCT ATG GAA GGC AAG AGG TCG CCG
 1: ser asn tyr phe ser met asp ser met glu gly lys arg ser pro

321 341
 TAC GCA GGG CTC CAG CTG GGG GCT GCC AAG AAG CCA CCC GTT ACC
 1: tyr ala gly leu gln leu gly ala ala lys lys pro pro val thr

361 381 401
 TTT GCC GAA AAG GGC GAC GTG CGC AAG TCC ATT TTC TCG GAG TCC
 1: phe ala glu lys gly asp val arg lys ser ile phe ser glu ser

421 441
 CGG AAG CCC ACG GTG TCC ATC ATG GAG CCC GGG GAG ACC CGG CGG
 1: arg lys pro thr val ser ile met glu pro gly glu thr arg arg

FIG. 5A.

8/21

				461										481					
	AAG	AGC	TAC	CCC	GGG	GCC	GAC	ACG	GGC	CTT	TTT	TCA	CGG	TCC	AAG				
1:	asn	ser	tyr	pro	arg	ala	asp	thr	gly	leu	phe	ser	arg	ser	lys				
				501										521					
	TCC	GGC	TCC	GAG	GAG	GTG	CTG	TGC	GAC	TCC	TGC	ATC	GGC	AAC	AAG				
1:	ser	gly	ser	glu	glu	val	leu	cys	asp	ser	cys	ile	gly	asn	lys				
				541										561					581
	CAG	AAG	GCG	GTC	AAG	TCC	TGC	CTG	GTG	TGC	CAG	GCC	TCC	TTC	TGC				
1:	gln	lys	ala	val	lys	ser	cys	leu	val	cys	gln	ala	ser	phe	cys				
							601							621					
	GAG	CTG	CAT	CTC	AAG	CCC	CAC	CTG	GAG	GGC	GCC	GCC	TTC	CGA	GAC				
1:	glu	leu	his	leu	lys	pro	his	leu	glu	gly	ala	ala	phe	arg	asp				
							641							661					
	CAC	CAG	CTG	CTC	GAG	CCC	ATC	CGG	GAC	TTT	GAG	GCC	CGC	AAG	TGT				
1:	his	gln	leu	leu	glu	pro	ile	arg	asp	phe	glu	ala	arg	lys	cys				
							681							701					
	CCC	GTG	CAT	GGC	AAG	ACG	ATG	GAG	CTC	TTC	TGC	CAG	ACC	GAC	CAG				
1:	pro	val	his	gly	lys	thr	met	glu	leu	phe	cys	gln	thr	asp	gln				
														721					741
	ACC	TGC	ATC	TGC	TAC	CTT	TGC	ATG	TTC	CAG	GAG	CAC	AAG	AAT	CAT				
1:	thr	cys	ile	cys	tyr	leu	cys	met	phe	gln	glu	his	lys	asn	his				
																			761
							781							801					
	AGC	ACC	GTG	ACA	GTG	GAG	GAG	GCC	AAG	GCC	GAG	AAG	GAG	ACG	GAG				
1:	ser	thr	val	thr	val	glu	glu	ala	lys	ala	glu	lys	glu	thr	glu				
																			821
	CTG	TCA	CTG	CAA	AAG	GAG	CAG	CTG	CAG	CTC	AAG	ATC	ATT	GAG	ATT				
1:	leu	ser	leu	gln	lys	glu	gln	leu	gln	leu	lys	ile	ile	glu	ile				
																			841
																			861
	GAG	GAT	GAA	GCT	GAG	AAG	TGG	CAG	AAG	GAG	AAG	GAC	CGC	ATC	AAG				
1:	glu	asp	glu	ala	glu	lys	trp	gln	lys	glu	lys	asp	arg	ile	lys				
																			881

FIG. 5B.

9/21

901
 AGC TTC ACC ACC AAT GAG AAG GCC ATC CTG GAG CAG AAC TTC CGG
 1: ser phe thr thr asn glu lys ala ile leu glu gln asn phe arg

921
 941
 GAG CTG GTG CGG GAC CTG GAG AAG CAA AAG GAG GAA GTG AGG GCT
 1: asp leu val arg asp leu glu lys gln lys glu glu val arg ala

961
 981
 GCG CTG GAG CAG CGG GAG CAG GAT GCT GTG GAC CAA GTG AAG GTG
 1: ala leu glu gln arg glu gln asp ala val asp gln val lys val

1001
 1021
 ATC ATG GAT GCT CTG GAT GAG AGA GCC AAG GTG CTG CAT GAG GAC
 1: ile met asp ala leu asp glu arg ala lys val leu his glu asp

1041
 1061
 AAG CAG ACC CGG GAG CAG CTG CAT AGC ATC AGC GAC TCT GTG TTG
 1: lys gln thr arg glu gln leu his ser ile ser asp ser val leu

1081
 1101
 1121
 TTT CTG CAG GAA TTT GGT GCA TTG ATG AGC AAT TAC TCT CTC CCC
 1: phe leu gln glu phe gly ala leu met ser asn tyr ser leu pro

1141
 1161
 CCA CCC CTG CCC ACC TAT CAT GTC CTG CTG GAG GGG GAG GGC CTG
 1: pro pro leu pro thr tyr his val leu leu glu gly glu gly leu

1181
 1201
 GGA CAG TCA CTA GGC AAC TTC AAG GAC GAC CTG CTC AAT GTA TGC
 1: gly gln ser leu gly asn phe lys asp asp leu leu asn val cys

1221
 1241
 1261
 1281
 1301
 ATG CGC CAC GTT GAG AAG ATG TGC AAG GCG GAC CTG AGC CGT AAC
 1: met arg his val glu lys met cys lys ala asp leu ser arg asn

1321
 1341
 TTC ATT GAG AGG AAC CAC ATG GAG AAC GGT GGT GAC CAT GGC TAT
 1: phe ile glu arg asn his met glu asn gly gly asp his arg tyr

FIG. 5C.

10/21

1361										1381																			
GTG AAC AAC TAC ACG AAC AGC TTC GGG GGT GAG TGG AGT GCA CCG																													
1: val asn asn tyr thr asn ser phe gly gly glu trp ser ala pro																													
1401										1421																			
GAC ACC ATG AAG AGA TAC TCG ATG TAC CTG ACA CCC AAA GGT GGG																													
1: asp thr met lys arg tyr ser met tyr leu thr pro lys gly gly																													
1441										1461										1481									
GTC CGG ACA TCA TAC CAG CCC TCG TCT CCT GGC CGC TTC ACC AAG																													
1: val arg thr ser tyr gln pro ser ser pro gly arg phe thr lys																													
1501										1521																			
GAG ACC ACC CAG AAG AAT TTC AAC AAT CTC TAT GGC ACC AAA GGT																													
1: glu thr thr gln lys asn phe asn asn leu tyr gly thr lys gly																													
1541										1561																			
AAC TAC ACC TCC CGG GTC TGG GAG TAC TCC TCC AGC ATT CAG AAC																													
1: asn tyr thr ser arg val trp glu tyr ser ser ser ile gln asn																													
1581										1601																			
TCT GAC AAT GAC CTG CCC GTC GTC CAA GGC AGC TCC TCC TTC TCC																													
1: ser asp asn asp leu pro val val gln gly ser ser ser phe ser																													
1621										1641										1661									
CTG AAA GGC TAT CCC TCC CTC ATG CGG AGC CAA AGC CCC AAG GCC																													
1: leu lys gly tyr pro ser leu met arg ser gln ser pro lys ala																													
1681										1701																			
CAG CCC CAG ACT TGG AAA TCT GGC AAG CAG ACT ATG CTG TCT CAC																													
1: gln pro gln thr trp lys ser gly lys gln thr met leu ser his																													
1721										1741																			
TAC CGG CCA TTC TAC GTC AAC AAA GGC AAC GGG ATT GGG TCC AAC																													
1: tyr arg pro phe tyr val asn lys gly asn gly ile gly ser asn																													
1761																													
GAA GCC CCA TGA																													
1: glu ala pro ***																													

FIG. 5D.

11/21

1Kb

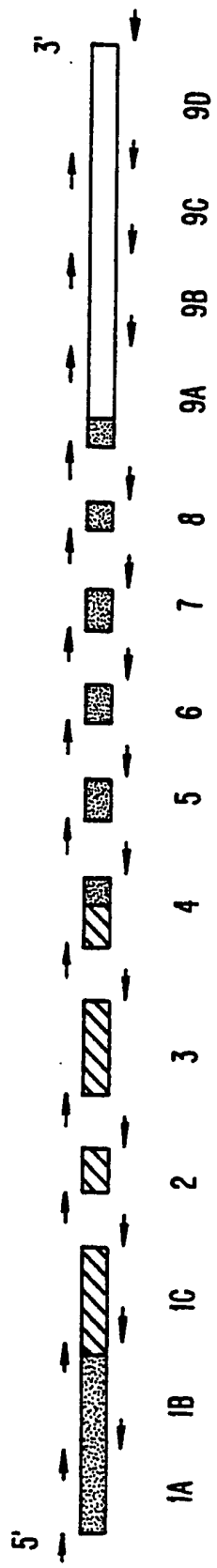


FIG. 6.

12/21

FRAGMENT	PRIMER	SEQUENCE
1A	1A-5	TCTCTAGTCCTCGTGGTT
	1A-3	GGTGGTCTTGGCATCCTT
1B	1B-5	TGGAGAATGGCACCAAGG
	1B-3	TCCATGATGGACACGGTG
1C	1C-5	TCTATGGAAGGCAAGAGG
	1C-3	GGAGAAGATGAAATTGCG
2	2-5	TGACTTCTCCAATCCTGG
	2-3	CCTGGACTCAAATGGGAG
3	3-5	AAGACATACCCGACTAGG
	3-3	TGTGAAATCGAGGGCTTG
4	4-5	AGCGTCCTCATAGCTCAT
	4-3	TGAGAAGAAGCTCACTGG
5	5-5	AAACTTGATCTGCCTGG
	5-3	AGTCACTGCACGGACTTT
6	6-5	GAGTCCTGATGAGACAAT
	6-3	CATTTCATCTCACACTGGG
7	7-5	AGAGAGTCATAGACCTGG
	7-3	GAGGAACTAGCAGCTCAG
8	8-5	GACGGCTGCATTTGGTAA
	8-3	CAGAGAAGTCCTCCACA
9A	9A-5	AGAATTGTCGGGTCTTGG
	9A-3	GCACAGTAGCTTAGAAGG
9B	9B-5	ACAAAGGCAACGGGATTG
	9B-3	TCTGCTGACAACATTGGG
9C	9C-5	AGACGTTTCTCTCCAAGG
	9C-3	CTTTATCTCAGCAGAGGG
9D	9D-5	AGGATGACCTTAGCCAAG
	9D-3	GAAGAAGTGCAGCCTGTT

FIG. 7.

13/21

1A
(SEQUENCED STRAND)

1A-5

1 TCTCTAGTCCTCGTGGTTGCCTGCCCCACTCCCTGCCGAG
41 AGGCCTGCCAGAAAGGTCACCTATCCTGAACCCCAAGCAAG
81 CCTGAAACAGCTCAGCCAAGCACCTGCGATGGAAGCTGC
121 AGATGCCTCCAGGAGCAACGGGTCGAGCCCAGAAAGCCAGG
161 GATGCCCGGAGCCCGTCGGGCCCCAGTGGCAGCCTGGAGA
201 ATGGCACCAAGGCTGACGGCAAGGATGCCAAGACCACC

IA-3

FIG. 8A.

1B
(SEQUENCED STRAND)

1B-5

1 TGGAGAATGGCACCAAGGCTGACGGCAAGGATGCCAAGAC
41 CACCAACGGGCACGGCGGGGAGGCAGCTGAGGGCAAGAGC
81 CTGGGCAGCGCCCTGAAGCCAAGGGAAGGTAGGAGCGCCC
121 TGTTCCGGGGCAATGAGTGGCGGGGACCCATCATCCAGTT
161 TGTGAGTCCGGGGACGACAAGAACTCCAACACTTCAGC
201 ATGGAATCTATGGAAGGCAAGAGGTGCGCGTACGCAGGGC
241 TCCAGCTGGGGGCTGCCAAGAAGCCACCCGTTACCTTTGC
281 CGAAAAGGGCGACGTGCGCAAGTCCATTTCTCGGAGTCC
321 CGGAAGCCACGGTGTCCATCATGGA

1B-3

FIG. 8B.

14/21

1C
(SEQUENCED STRAND)

1C-5

1 TCTATGGAAGGCAAGAGGTCGCCGTACGCAGGGCTCCAGC
41 TGGGGGCTGCCAAGAAGCCACCCGTTACCTTTGCCGAAAA
81 GGGCGACGTGCGCAAGTCCATTTTCTCGGAGTCCCGGAAG
121 CCCACGGTGTCCATCATGGAG CCCGGGGAGACCCGGCGGA
161 ACAGCTACCC CCCGGGCGGACACGGGCCTTTTTTCACGGTC
201 CAAGTCCGGCTCCGAGGAGGTGCTGTGCGACTCCTGCATC
241 GGCAACAAGCAGAAGGCGGTCAAATCCTGCGTGGTGCC
281 AGGCCTCCTTCTGCGAGCTGCATCTCAAGCCCGACCTGGA
321 GGGCGCGCCTTCCGAGACCACCAAGTGC CTCGAGCCCATC
361 CGGGACTTTGAGGCCCGCAAGTGTCCCGTG CATGGCAAGA
401 CGATG GAGCTCTTCTGCCAGACCGACAGACCTGCATCTG
441 CTACCTTTGCATGTTCCAGGAGCACAAGAATCATAGCACC
481 GTGACAGTGGAGGAGGCCAAGGCCGAGAAGGAG GTAAGTG
521 CTGGGGCCCTCCTGCCCTCCAGGCCTCTCCTCTCTCAA
561 CCCACCCCT CCGAAGTTCATCTTCTCC

1C-3

FIG. 8C.

15/21

2

(SEQUENCED STRAND)

1 2-5 INTRON, EXON2
 TGACTTCTCCAATCCTGGCTCTTTCTCTGCAGACGGAGCT
 41 Pst I Pst I
 GTCAGTGC AAAAGGAGCAGCTGCAGCTCAAGATCATTGAG
 81 ATTGAGGATGAAGCTGAGAAGTGGCAGAAGGAGAAGGACC
 EXON 2 INTRON
 121 GCATCAAGGTGAGCAGCCCCAAGCTCACCTTGCTGCTCC
 161 CTTACCCGACCTGGCCTGCCTGGAAAGACGCAAGCCTTG
 201 CTCCCATTTGAGTCCAGG
 2-3

FIG. 8D.

3

(SEQUENCED STRAND)

1 3-5
 AAGACATACCCGACTAGGGTGATTTCCTTCCCTAACTAAA
 41 INTRON, EXON3
 GCCCTGCCTAATCTCTTCCCTGACTCTGGACCTCCAGAGC
 81 TTCACCACCAATGAGAAGGCCATCCTGGAGCAGAACTTCC
 121 GGGACCTGGTGGGGACCTGGAGAAGCAAAAGGAGGAAGT
 161 GAGGGCTGGCTGGAGCAGCAGGAGCAGGATGCTGTGGAC
 201 CAAGTGAAGGTGATCATGGATGCTCTGGATGAGAGAGCCA
 241 AGGTGCTGCATGAGGACAAGCAGACCCGGGAGCAGCTGCA
 EXON3, INTRON
 281 TAGCATCAGCGACTCTGTGTTGTTTCTGCAGGTAACAAGC
 321 CACTCCTCTGTCACTCAAGCCCTCGATTTCACA
 3-3

FIG. 8E.

16/21

4
(SEQUENCED STRAND)

4-5
1 AGCGTCCTCATAGCTCATGAAGACCCAGGCAGTTAATGGT
INTRON, EXON 4
41 TCTTTCCTTTCTTGGTAGGAATTTGGTGATTGATGAGCA
81 ATTACTCTCTCCCCCACCCTGCCACCTATCATGTCCT
121 GCTGGAGGGGAGGGCCTGGGACAGTCACTAGGCAACTTC
161 AAGGACGACCTGCTCAATGTATGCATGCGNCACGTTGAGA
201 AGATGTGCAAGGCGNACCTGAGCNGTAACTTCATTGAGAG
EXON 4, INTRON
241 GAACCACATGGAGAACG|GTAGGTCCCCTCTCGTGGCTGGG
281 CCCCAGGCNATAGACGTTTCTCTCCAAATCAATTCCCTG
321 CTGCCTGACATGGGCTGGCCTCCAGTGAGCTTCTTCTCA

4-3

FIG. 8F.

5
(SEQUENCED STRAND)

5-5
1 AAACTTGGATCTGCCTGGGAGATAGGGGAAGGGCTATGGG
41 GTGACTCATCTGAGCCCCAAAAGTCCCAGTGGCTGGCTC
81 CTCCTTCCACCTGGCTCCTCTGCTGACCCGACCCTCTGC
INTRON, EXON 5
121 TTCCTAGGTGGTGACCATCGCTATGTGAACAACTACACGA
161 ACAGCTTCGGGGGTGAGTGGAGTGACCCGGACACCATGAA
EXON 5, INTRON
201 GAGATACTCCATGTACCTGACACCCAAAG|GTAAGAGGGAG
241 CCCCTCACCCAGACCTAGTGTCTCTCCTGCTGCCAGGG
281 GCCCCCAAAGTCCGTGCAGTGA

5-3

FIG. 8G.

17/21

6
(SEQUENCED STRAND)

6-5
1 GAGTCCTGATGAGACAATTTTGTGCAATGACAGCCCNNTT
INTRON EXON6
41 CATCTGCTTCACAGGTGGGGTCCGGACATCATACCAGCCC
81 TCGTCTCCTGGCCGCTTCACCAAGGAGACCACCAAGAAGA
EXON6 INTRON
121 ATTTCAACAATCTCTATGGCAGCAAAGNNNNNNNTGGGNC
161 TGTGCAGGCAAGAGGGCATAGAGGTGGGTCCAGNGGCACA
201 GGGCTGGGACCCAGTGTGAGATGAATG
6-3

FIG. 8H.

7
(SEQUENCED STRAND)

7-5
1 AGAGAGTCATAGACCTGGCTGTGTCCTGGTCCTGCCTCCT
41 CTCCCACTCCCAGCTGTGGGGGCTGACAGCCCTTCTTTG
INTRON EXON7 SmaI
81 TCCTGCCAGGTAACCTACACCTCCCGGGTCTGGGAGTACTC
121 CTCCAGCATTGAGAACTCTGACAATGACCTGCTGTGCTCC
EXON7 INTRON
161 AAGGCAGCTCCTCCTTCTCCCTGAAAGGTGAGCCCTGCCC
201 ACCCTGGCCCCTGCTTTCTCCACAGCTGCCTCACACCTC
241 CCAAGCCCTGCTTGGGTCTCTTCGCTGAGCTGCTAGTTCC
7-3
281 TC

FIG. 8I.

18/21

8
(SEQUENCED STRAND)

8-5

1 GACGGCTGCATTTGGTAATGGGCTGGATGATGCTTGGTGG

41 TACACTTTGGAGAAGNAGCTGTGCTGCTCTGGGNCCGGGN

81 NCCCCTGGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
INTRON | EXON 8

121 NNNNNNNNNNNNGCTATCCCTCCCTCATGCGGAGCCAAAG

161 CCCCAGGCCAGCCCCAGACTTGGAAATCTGGCAAGCAG

201 EXON 8 | INTRON
ACTATGCTGTAAAGGAAGTGGGCCG66GAGGGCCTGGGC

241 ACATCCAGAGACCTGGGCACTGAAGGGGGCTCCCTGGAGG

281 CAATCGGTTCCAGGGCCTGTGGGAGGACTTCTCTG

8-3

FIG. 8J.

9A
(UNSEQUENCED STRAND)

9A-5

1 AGAATTGTCGGGTCTTGGATCACTGCTGCCCTCTGAGGCA

41 GGTAGGGTAGGGTGGGTCTAGCTAGCAGGCTCATCTGTC

81 GTCTGGCCTCGCTGACCACTCTTGTTCCTCCACAGTCTC
INTRON | EXON 9

121 ACTACGGGCATTCTACGTCAACAAAGGCAACGGGATTGG

161 STOP SstI
GTCCAACGAAGCCCCATGAGCTCCTGGCGGAAGGAACGAG

201 GCGCCACACCCCTGCTCTTCCTCCTGACCCTGCTGCTCTT

241 GCCTTCTAAGCTACTGTGC

9A-3

FIG. 8K.

19/21

9B
(SEQUENCED STRAND)

9B-5

1 ACAAAGGCAACGGGATTGGGTCCAACGAAGCCCCATGAGC

41 TCCTGGCGGAAGGAACGAGGGGCCACACCCCTGCTCTTCC

81 TCCTGACCCTGCTGCTCTTGCCTTCTAAGCTACTGTGCTT

121 GYCTGGGTGGGAGGGAGCCTGGTCCTGCACCTGCCCTCTG^{PstI}

161 CAGCCCTCTGCCAGCCTCTTGGGGGCAATTCCGGCCTCTC

201 CGACTTCCCCACTGGCCACACTCCATTGAGACTCCTTTCC

241 TGCTTGTGACCTCAGATGGTCACCATCATTCTGTGCTC

281 AGAGGCCAACCCTATCACAGGGGTGAGATAGGTTGGGGCCT

321 GCCCTAACCGGCCAGCCTCCTCCTCTCGGGCTGGATCTGG

361 GGGCTAGCAGTGAGTACCGGCATGGTATCAGCCTGCCTCT

401 CCGGCCACGCCCTGCTGTCTCCAGGCCTATAGACGTTTC

441 TCTCCAAGGCCCTATCCCCCAATGTTGTCAGCAGA

9B-3

FIG. 8L

20/21

9C

(SEQUENCED STRAND)

9C-5

1 AGACGTTTCTCTCCAAGGCCCTATCCCCAATGTTGTCAG

41 CAGATGCCTGGACAGCACAGCCACCCATCTCCCATTGACA

81 TGGCCCAGCTCCTGCTTCCCAGAGGACTGGCCCTACGTGC

121 TCTCTCTCGTCCTACCTATCAATGCCCAGCATGGCAGAAC

Pst I Pst I

161 CTGCAGTGGCCAAGGGCTGCAGATGGAAACCTCTCAGTGT

201 CTTGACATCACCTACCCAGGGGGTGGGTCTCCACCACAG

241 CCACTTTGAGTCTGTGGTCCCTGGAGGGTGGCTTCTCCTG

281 ACTGGCAGGATGACCTTAGCCAAGATATTCCTCTGTTCCC

321 TCTGCTGAGATAAAG

9C-3

FIG. 8M.

9D
(SEQUENCED STRAND)

9D-5

1 AGGATGACCTTAGCCAAGATATTCCTCTGTTCCCTCTGCT

41 GAGATAAAGAATTCCCTTAACATGATATAATCCACCCATG
EcoRI

81 CAAATAGCTACTGGCCCAGCTACCATTACCATTTCCTA

121 CAGAATTTCAATTCAGTCTACACTTTGGCATTCTCTCTGGC

151 GATGGAGTGTTGGCTGGGCTGACCGCAAAGGTGCTTACA

201 CACTGCCCCCACCCTCAGCCGTTGCCCCATCAGAGGCTGC

241 CTCCTCCTTCTGATTACCCCCCATGTTGCATATCAGGGTG

281 CTC AAGGATTGGAGAGGAGACAAAACCAGGAGCAGCACAG

321 TGGGGACATCTCCCGTCTCAACAGCCCCAGGCCTATGGGG

361 GCTCTGGAAGGATGGGCCAGCTTG CAGGGGTTGGGGAGGG

401 AGACATCCAGCTTGGGCTTTCCCTTTGGAATAAAC CATT

441 GGTCTGTC ACTTCTCTTG TATTGAATGACCATTTCCTGA
cDNA - END

481 GGGTCCC CAGAGGAACAGGCTG CAGTTCTTC

9D-3

FIG. 8N.